

RUMOURS SPREAD SLOWLY IN A SMALL WORLD SPATIAL NETWORK

JEANNETTE JANSSEN AND ABBAS MEHRABIAN

ABSTRACT. Rumour spreading is a protocol for modelling the spread of information through a network via user-to-user interaction. The Spatial Preferred Attachment (SPA) model is a random graph model for complex networks: vertices are placed in a metric space, and the link probability depends on the metric distance between vertices, and on their degree. We show that the SPA model typically produces graphs that have small effective diameter, i.e. $O(\log^2 n)$, while rumour spreading is relatively slow, namely polynomial in n .

1. INTRODUCTION

There is increasing consensus in the scientific community that complex networks (e.g. on-line social networks or citation graphs) can be accurately modelled by spatial random graph models. Spatial random graph models are models where the vertices are located in a metric space, and links are more likely to occur between vertices that are close together in this space. The space can be interpreted as a *feature space*, which models the underlying characteristics of the entities represented by the vertices. Specifically, entities with similar characteristics (for example, users in a social network that share similar interests) will be placed close together in the feature space. Thus the distance between vertices is a measure of affinity, and thus affects the likelihood of the occurrence of a link between these vertices.

An important reason to model real-life networks is to be able, through simulation or theoretical analysis, to study the dynamics of information flow through the network. Several ways to model flow of information through a network have been proposed recently, based on metaphors such as the spread of infection or of fire, or the range of a random walk through the graph [4, 7, 22, 25]. Here we focus on a protocol called *rumour spreading*. It differs from the models based on fire or infection in that in each round, the rumour spreads to only one neighbour of each informed vertex. On the other hand, the difference with a random

Date: August 8, 2016.

2010 Mathematics Subject Classification. 05C82; 05C80, 60D05, 90B15, 91D30.

Key words and phrases. Spatial Preferred Attachment model; rumour spreading; effective diameter; push&pull protocol.

The collaboration between the authors was a result of the visit of the first author to the Institute of Mathematics and Applications (IMA) in Minnesota. She wishes to thank IMA for providing this opportunity. She also acknowledges NSERC for their support of this research.

The second author was supported by the Vanier Canada Graduate Scholarships program, a PIMS Postdoctoral Fellowship, and an NSERC Postdoctoral Fellowship.

A preliminary version of this paper appeared in proceedings of the 12th workshop on algorithms and models for the web graph, WAW 2015, published by Springer in the Lecture Notes for Computer Science 9479, pp. 107–118. In this full version, all the proofs are included, and the upper bound for the effective diameter is improved to $O(\log^2 n)$.

walk approach is that each informed vertex spreads the rumour, and thus we have more of a growing tree of random walks.

In this paper we study the behaviour of the rumour spreading protocol on graphs produced by the Spatial Preferential Attachment (SPA) model, which is a spatial model that produces sparse power law graphs. We show that, on the one hand, the graph distance between vertices in such a graph tends to be small (polylogarithmic in n , the number of vertices), while on the other hand, it takes a long time (polynomial in n) to spread the rumour to most of the vertices.

1.1. The SPA model. The SPA model is a growing graph model, where one new vertex is added to the graph in each time step. The vertices are chosen from a metric space. Each vertex has a sphere of influence, whose size grows with the degree of the vertex. A new vertex can only link to an existing vertex if it falls inside its sphere of influence. Therefore, links between vertices depend on their (spatial) distance, and on the in-degree of the older vertex.

The SPA model was introduced in [1], where it was shown that asymptotically, graphs produced by the SPA model have a power law degree distribution with exponent in $[2, \infty)$ depending on the parameters. The model was further studied in [8, 19, 20]. The model can be seen as a special case of the spatial model introduced by Jacob and Mörters in [18] and further studied in [17]. The SPA model has similarities with the spatial models introduced in [3, 5, 12, 27].

Let S be the unit hypercube in \mathbb{R}^m , equipped with the torus metric derived from the Euclidean norm. The SPA model stochastically generates a graph sequence $\{G_t\}_{t \geq 0}$; for each $t \geq 0$, $G_t = (V_t, E_t)$, where E_t is an edge set, and $V_t \subseteq S$ is a vertex set. The index t is an indication of time. The in-degree, out-degree and total degree of a vertex v at time t is denoted by $\deg^-(v, t)$, $\deg^+(v, t)$ and $\deg(v, t)$, respectively.

We now define the *sphere of influence* $S(v, t)$ of a vertex v at time t . Let

$$A(v, t) := \frac{A_1 \deg^-(v, t) + A_2}{t},$$

where $A_1, A_2 > 0$ are given parameters. If $A(v, t) \leq 1$, then $S(v, t)$ is defined as the ball, centred at v , with total volume $A(v, t)$. If $A(v, t) > 1$ then $S(v, t) = S$, and so $|S(v, t)| = 1$. To keep the second option from happening often, we impose the additional restriction that $A_1 < 1$; this ensures that in the long run, $S(v, t) \ll 1$ for all v .

The generation of a SPA model graph begins at time $t = 0$ with G_0 being the null graph. At each time step $t \geq 1$, a node v_t is chosen from S according to the uniform distribution, and added to V_{t-1} to form V_t . Next, independently for each vertex $u \in V_{t-1}$ such that $v_t \in S(u, t)$, a directed link (v_t, u) is created with probability p .

Because the volume of the sphere of influence of a vertex is proportional to its in-degree, so is the probability of the vertex receiving a new link at a given time. Thus link formation is governed by a preferential attachment, or “rich get richer”, principle, which leads to a power law degree distribution of the in-degrees, and thus also of sizes of the spheres of influence.

Another important feature of the model is that all spheres of influence tend to shrink over time. This means that the length of an edge (the distance between its endpoints) depends on the time when it was formed: edges formed in the beginning of the process tend to be much longer than those formed later (see [19] for more on the distribution of edge lengths).

As we will see, the old, long links significantly decrease the graph distance between vertices. This is a feature unique to the SPA model; “static” variations of the SPA model such as that presented in [2], tend to limit the maximum length of an edge, which leads to a larger diameter.

Note that the SPA model generates directed graphs. However, the rumour spreading protocols we study here completely ignore the edge orientations; we imagine that they work on the corresponding undirected underlying graph. Similarly, in estimating the graph distances, we ignore the edge orientations.

1.2. Rumour spreading. Rumour spreading is a model for the spread of one piece of information, the *rumour*, which starts at one vertex, and in each time step, spreads along the edges of the graph according to one of the following protocols.

The *push protocol* is a round-robin rumour spreading protocol defined as follows: initially one vertex of a simple undirected graph knows a rumour and wants to spread it to all other vertices. In each round, every informed vertex sends the rumour to a random neighbour.

The *push&pull protocol* is another round-robin rumour spreading protocol defined as follows: initially one vertex of a simple undirected graph knows a rumour and wants to spread it to all other vertices. In each round, every informed vertex sends the rumour to a random neighbour, while every uninformed vertex contacts a random neighbour and gets the rumour from her if she knows it.

In both protocols defined above, all vertices work in parallel. These are synchronized protocols, so if a vertex receives the rumour at round t , it starts passing it on from round $t + 1$. Also, vertices do not have memory, so a vertex might contact the same neighbour in consecutive rounds.

We are interested in the *spread time*, the number of rounds needed for all vertices to get informed. Since the SPA model does not generally produce connected graphs, we here limit this requirement to vertices in the same component as the starting vertex. It is clear that the push&pull protocol is generally quicker (this can be made precise via a coupling argument).

The push protocol was defined in [15] for the complete graph, and was studied in [11] for general graphs. The push&pull protocol was defined in [9], where experimental results were presented, and the first analytical results appeared in [21].

1.3. Main results. Clearly, the diameter of a graph is a lower bound on the spread time, at least for appropriate choices of starting vertex. An easy well known upper bound for spread time is $O(\Delta(\text{diameter} + \log n))$ [11, Theorem 2.2], where Δ denotes the maximum degree. So in graphs of bounded degree, spread time is largely determined by the diameter. Another important factor in rumour spreading is the degree distribution of the graph. Vertices of high degree tend to slow down the spread, since only one neighbour of a vertex is contacted in each round. SPA model graphs have a power law degree distribution, and the maximum degree is typically $\Omega(n^{A_1})$ (see [1]).

In this paper we prove two main results. First, we show that for most pairs of vertices, the graph distance is polylogarithmic in the number of vertices. Thus, SPA model graphs are so-called small worlds. SPA model graphs are generally not connected, and the size and threshold of the giant component are not exactly known. Therefore we state our result in terms of the *effective diameter*, introduced in [23]. A graph G has effective diameter at most d if, for at least 90% of all pairs of vertices of G that are connected, their graph distance is

at most d . We say an event happens *asymptotically almost surely* (a.a.s.) if its probability approaches 1 as n goes to infinity. All logarithms are in the natural base in this paper.

Recall that the SPA model has four parameters: $m \in \mathbb{Z}_+$ is the dimension, $A_1, A_2 > 0$ control the volumes of vertices' spheres of influence, and $p \in (0, 1]$ is the probability of link formation.

Theorem 1.1. *For each choice of $A_1 \in [0, 1)$, and for large enough choice of A_2 , a.a.s. a graph produced by the SPA model with parameters A_1, A_2 , $p = 1$ and $m = 2$ has effective diameter $O(\log^2 n)$.*

Remark. The constant 90% in the definition of effective diameter is somewhat arbitrary. Our arguments yield similar bounds if this is changed to any other constant $1 - \varepsilon$ strictly smaller than 100%.

As noted before, this result refers to the *undirected* diameter. In [8], it was shown that a.a.s. any shortest directed path has length $O(\log n)$. This result does not apply to our situation, since pairs connected by a directed path are a small minority.

Bringmann, Keusch, and Lengler [6] proved a polylogarithmic upper bound for the diameter of a spatial random graph model with given expected degrees. In their model the edges appear independently, and so their result does not apply to our model.

We believe the conclusion of Theorem 1.1 is not tight. We make the following conjecture.

Conjecture 1.2. *For each choice of $p, A_1 \in (0, 1)$, and for large enough choice of A_2 , a.a.s. the giant component of a graph produced by the SPA model with parameters A_1, A_2 , p and $m = 2$ has diameter $O(\log n)$.*

Our proof for Theorem 1.1 is based on two-dimensional objects called crossings, and for extending it to higher dimensions new techniques are required. We leave this as an open problem.

Question 1.3. *Extend Theorem 1.1 to higher dimensions $m > 2$.*

Our second result illustrates that, in spite of the small world property, a.a.s. rumour spreading with the push&pull protocol is slow, that is, takes polynomial time in n .

Theorem 1.4. *Let G be a graph produced by the SPA model with parameters $A_1, A_2 > 0, m \in \mathbb{Z}_+$, and assume that $a := pA_1 < 1$. Define $K := (3 + a)m + 1 - a$ and let $\alpha < a(1 - a)/K$ be a constant. If a rumour starts in G from a uniformly random vertex, then a.a.s. after n^α rounds of the push&pull protocol, the number of informed vertices is $o(n)$.*

Remark. Note that, since the push protocol is not quicker than the push&pull protocol, the same lower bound holds for the push protocol.

Let us also remark that our diameter result, Theorem 1.1, holds only for dimension 2, and this is because our argument is based on building crossings, whereas our rumour spreading bound holds for all dimensions.

We can understand Theorem 1.4 as follows. While SPA model graphs have a backbone of long edges that decrease graph distances between vertices, only old edges are long. Old edges have old endpoints, so the vertices on this backbone are old. Old vertices have high degree, and vertices of high degree are slower in spreading the rumour. So if the rumour

travels along long edges, then it will become delayed due to high vertex degree, and if it travels along short edges, it takes many steps to cover the entire space.

In [18] it was shown that, for certain choices of the parameters, the generalized spatial model by Jacob and Mörters exhibits a similar mixture of long and short edges. This suggests that our results may be extended to this model; this would be an interesting question to pursue.

Question 1.5. *Can our results be extended to the Jacob-Mörters model [18]?*

The push&pull protocol has been studied on two small-world (non-spatial) models and it turned out that it spreads the rumour in logarithmic time: it was shown in [10] that on a random graph model based on preferential attachment, push&pull spreads the rumour within $O(\log n)$ rounds. A similar bound was proved for the performance of this protocol on random graphs with given expected degrees when the average degree distribution is power law [13]. Thus, the SPA model is a unique example of a natural model that exhibits both the small world property and slow rumour spreading.

2. THE EFFECTIVE DIAMETER OF SPA MODEL GRAPHS

In this section we prove Theorem 1.1, which states that a two-dimensional SPA model graph with $p = 1$ typically has a small effective diameter. Assume that $m = 2$ and $p = 1$. We will derive our bound using properties of the random geometric graph model, especially those studied in [14, 16, 26].

A two-dimensional random geometric graph on N vertices with radius $r = r(N)$, denoted by $RGG(N, r)$, is generated as follows: N vertices are chosen independently and uniformly at random from the unit square S , and an edge is added between two vertices if and only if their Euclidean distance is at most r . To see how the geometric random graph model relates to the SPA model, let $\{G_t\}_{t=0}^n$ be a sequence of graphs produced by the SPA model. Our analysis is based on a sequence of subgraphs of G_t , which mimic the behaviour of the model when the in-degree does not influence the size of the sphere of influence.

For each t , define the graph R_t as a graph with vertex set $V(G_t)$ in which two vertices are adjacent if and only if their distance is at most $\sqrt{A_2/t\pi}$. Observe that R_t conforms to the random geometric graph model on t vertices with radius

$$r_t := \sqrt{\frac{A_2}{t\pi}}. \quad (1)$$

For all t , R_t is a subgraph of (the undirected underlying graph of) G_t . Namely, at all times from 1 to t , each sphere of influence has volume at least A_2/t , i.e. radius at least r_t . Therefore, if two vertices v_i and v_j , $1 \leq i < j \leq t$, have distance at most r_t , then at time j , when v_j is born, v_j will fall inside the sphere of influence of v_i , and a link $v_j v_i$ will be created. We will use the graphs R_t to bound the diameter of G_n .

As mentioned earlier, graphs produced by the SPA model are generally not connected. However, we can choose the parameters so that there exists a giant component, i.e. a component that contains an $\Omega(1)$ fraction of all vertices. Note that if R_n has a giant component, then so has G_n . Moreover, it is known (see [26]) that there exists a constant a_c so that, if $r = \sqrt{\frac{a}{\pi N}}$ with $a > a_c$, then *a.a.s.* $RGG(N, r)$ has a giant component, while if $a < a_c$ then *a.a.s.* it does not have one (note that a is simply the average degree). Experiments give a

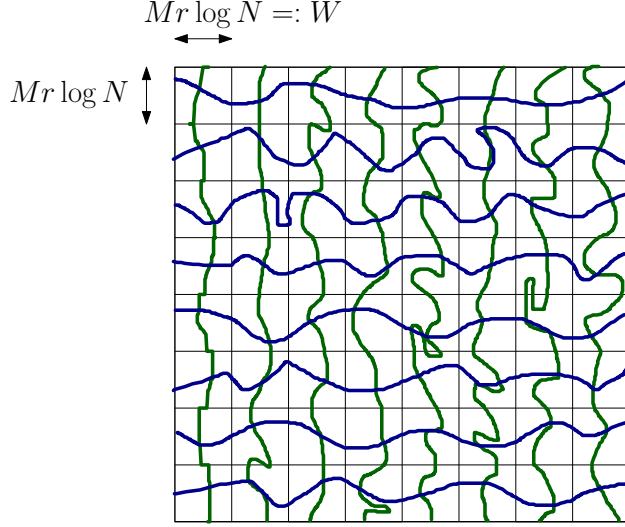


FIGURE 1. A spanning component: left-to-right crossings are blue, top-to-bottom crossings are green.

value of $a_c \approx 4.51$. Therefore, G_n has a giant component *a.a.s.* if $A_2 > a_c$. It would be interesting to determine whether this value of A_2 is indeed the threshold for the emergence of the giant component in G_n . Determination of this threshold was left as an open problem in [8, Section 5].

To show that the effective diameter of the giant component of G_n is $O(\log^2 n)$, we will proceed as follows. Given an arbitrary vertex v in the giant component of R_n , and thus of G_n , the idea is to find a path of length $O(\log n)$ connecting v to some vertex y_1 in the giant component of $R_{n/2}$, then connect y_1 to a vertex y_2 in the giant of $R_{n/4}$, and so on. It will be more convenient to work not with the giant components but with the so-called *spanning components*, which are defined next.

Definition 2.1 (Spanning component). *Let R be a random geometric graph with parameters N and r , and let $M := \pi Nr^2$. Let $W = W(N) := Mr \log N$, and assume that W^{-1} is an integer. Partition the unit square S into W^{-1} horizontal rectangles of size $W \times 1$ (the horizontal slabs), and also into W^{-1} vertical rectangles of size $1 \times W$ (the vertical slabs). For a horizontal slab L , a left-to-right crossing is a path $v_0 v_1 \dots v_k$ contained in L such that v_0 has distance $\leq r/5$ to the left side of L , v_k has distance $\leq r/5$ to the right side of L , and the distance between v_i and v_{i+1} is $\leq r/2$ for each $0 \leq i \leq k-1$. A top-to-bottom crossing is defined similarly for vertical slabs. It is easy to see that if R has a left-to-right crossing for each of the W^{-1} horizontal slabs, and has a top-to-bottom crossing for each of the W^{-1} vertical slabs, then the vertices of these crossings are contained in the same connected component, which is called a spanning component.*

See Figure 1 for an illustration. The following lemma guarantees the existence of spanning components.

Lemma 2.2. *There exists an absolute constant a_{large} such that a random geometric graph $R = \text{RGG}(N, r)$ with $\pi Nr^2 \geq a_{\text{large}}$ has a spanning component with probability $\geq 1 - O(N^{-1})$.*

Proof. Assume that the unit square S is subdivided into $r/5 \times r/5$ subsquares. Two subsquares are called *adjacent* if they share a side. A *path of subsquares* is a sequence of distinct subsquares so that each consecutive pair is adjacent. A subsquare is called *occupied* if it contains a vertex of R , and *empty* otherwise. For a given rectangle L that is tiled perfectly by subsquares, a *left-to-right subsquare crossing* is a path of occupied subsquares inside L with one endpoint touching the left side of L and the other endpoint touching its right side. A *top-to-bottom subsquare crossing* is defined similarly. Note that a left-to-right subsquare crossing indeed gives a left-to-right crossing as defined in Definition 2.1, by considering the vertices inside the corresponding subsquares (here we used the fact that, if we have two points lying in side by side squares of side length $r/5$, the distance between the points is less than $r/2$).

Let $M := \pi N r^2$. Ganesan [16, Lemma 2] proved that there exists an absolute constant $c > 0$ such that each of the horizontal (vertical) slabs (defined in Definition 2.1) has a left-to-right (top-to-bottom) subsquare crossing with probability at least $1 - O(N^{1/2-cM^2})$. The total number of slabs is

$$2/(Mr \log N) = O(\sqrt{N}/\log N),$$

so if $M \geq a_{\text{large}} := 2/\sqrt{c}$, then by the union bound, with probability at least $1 - O(N^{-1})$ all horizontal (vertical) slabs have a left-to-right (top-to-bottom) crossing. \square

For two vertices u and v of a random geometric graph R , we denote their Euclidean distance and graph distance by $d_E(u, v)$ and $d_R(u, v)$, respectively. The following result was proved by Friedrich, Sauerwald, and Stauffer (see Theorem 3 and Remark 5 in [14]).

Theorem 2.3 ([14]). *There exist absolute constants Γ and η such that a random geometric graph $R = \text{RGG}(N, r)$ with $\pi N r^2 > 5$ satisfies the following property with probability at least $1 - O(N^{-1})$: for any two vertices u, v that are in the same connected component of R and have $d_E(u, v) \geq \Gamma(\log N)/(rN)$, we have*

$$d_R(u, v) \leq \eta d_E(u, v)/r.$$

Definition 2.4. *Let Γ and η be the constants in Theorem 2.3. We say that a random geometric graph $R = \text{RGG}(N, r)$ is nice if*

- (i) *R has a spanning component, and*
- (ii) *for any two vertices u, v of R that are connected by a path and have $d_E(u, v) \geq \Gamma \log N/(rN)$, we have $d_R(u, v) \leq \eta d_E(u, v)/r$.*

Lemma 2.5. *Suppose that $A_2 \geq \max\{5, a_{\text{large}}\}$. Let k be the smallest integer such that $n2^{-k} \leq \log n$. A.a.s. we have that all random geometric graphs $R_n, R_{n/2}, \dots, R_{n2^{-k}}$ are nice.*

Proof. By the definition of k we have $n2^{1-k} > \log n$ and so $2^k < 2n/\log n$. For each $j \in \{n, n/2, \dots, n2^{-k}\}$, by Lemma 2.2 and Theorem 2.3 and the union bound, R_j is nice with probability at least $1 - O(j^{-1})$. By the union bound, the probability that at least one of these is not nice is bounded from above by

$$\sum_{i=0}^k (n/2^i)^{-1} = \frac{1}{n} \sum_{i=0}^k 2^i = \frac{2^k - 1}{n} < 2/\log n = o(1),$$

as required. \square

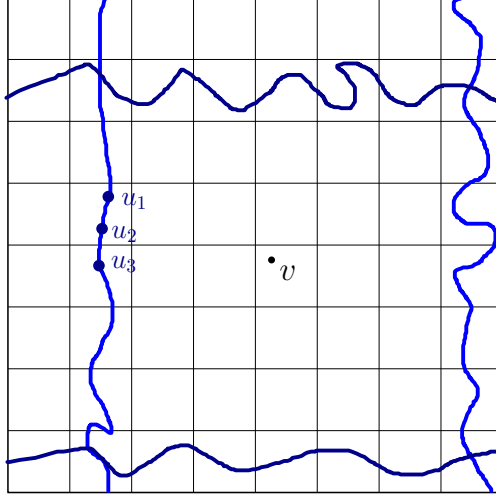


FIGURE 2. Illustration for the proof of Lemma 2.6, with $k = 3$: the crossings constitute a cycle $u_1 u_2 \dots$ enclosing the vertex v .

The following lemma is a purely deterministic one, and is the geometric core of our argument.

Lemma 2.6. *Suppose that $(\log n)/2 \leq t \leq n/2$ and that R_t and R_{2t} are nice. Let v be a vertex in the spanning component of R_{2t} . Then there exists a path of length $O(\log t)$ in R_{2t} from v to the spanning component of R_t .*

Proof. Let $M := \pi t r_t^2$, $W := M r_t \log t$, and consider the slabs of size $1 \times W$ and $W \times 1$ defined in Definition 2.1. Since R_t is nice, it has a spanning component, so each horizontal (vertical) slab has a left-to-right (top-to-bottom) subsquare crossing. Enumerate the slabs from left to right, and from top to bottom. Suppose v lies in the a -th horizontal slab and the b -th vertical slab.

Let $k := \lceil 2 + 3\Gamma/(\pi t^2 r_t^4) \rceil$. The left-to-right crossings of slabs $a - k$ and $a + k$ and the top-to-bottom crossings of slabs $b - k$ and $b + k$ constitute a cycle $u_1 u_2 \dots u_\ell$ in R_t , enclosing v , such that for all $i = 1, 2, \dots, \ell$, we have

- (i) $d_E(u_i, u_{i+1}) \leq r_t/2$ (by the definition of a crossing) and
- (ii) $(k - 1)W \leq d_E(v, u_i) \leq 2kW$.

See Figure 2.

We claim that there exists a path in R_{2t} from v to some u_i in the cycle. Let us show first that this claim completes the proof. Since $k > 2$, we have $k - 1 > k/2 > 3\Gamma/(2\pi t^2 r_t^4)$. By definition of W and M , we have $W = M r_t \log t = \pi t r_t^3 \log t$. Also, $r_{2t} = r_t/\sqrt{2}$ and $\log(2t) \leq 2 \log t$ for $t \geq 2$. Therefore,

$$\frac{\Gamma \log(2t)}{2tr_{2t}} \leq \frac{\sqrt{2}\Gamma \log t}{tr_t} < \frac{3\Gamma}{2\pi t^2 r_t^4} \times \pi t r_t^3 \log t < (k - 1)W.$$

Using property (ii) above, we find that

$$\Gamma \log(2t)/(2tr_{2t}) < (k - 1)W \leq d_E(v, u_i).$$

Since R_{2t} is nice, by property (ii) in Definition 2.4 we have that

$$d_{R_{2t}}(v, u_i) \leq \frac{\eta d_E(v, u_i)}{r_{2t}} \leq \frac{\eta 2kW}{r_{2t}} = O\left(\frac{kW}{r_t}\right) = O\left(\frac{k\pi tr_t^3 \log t}{r_t}\right) = O(ktr_t^2 \log t) = O(\log t),$$

proving the lemma. For the last equality we used the fact that $tr_t^2 = A_2/\pi = O(1)$ and also $k \leq 3 + 3\Gamma/\pi t^2 r_t^4 \leq 3 + 3\pi\Gamma/A_2^2 = O(1)$.

So it remains to show that, there exists a path in R_{2t} from v to some u_i in the cycle. Note that

$$W = Mr_t \log t = \pi tr_t^3 \log t = \pi t \log t \left(\frac{A_2}{t\pi}\right)^{3/2} = o(1),$$

and since $k = O(1)$ as we saw above, we find that $kW = o(1)$. Since all vertices of the cycle lie within distance $2kW = o(1)$ of v , and since v lies in a spanning component of R_{2t} , there is a path ζ in R_{2t} from v to a vertex not enclosed by the cycle. Let x_1 be the last vertex in ζ enclosed by the cycle, and let x_2 be the first vertex in ζ not enclosed by the cycle. If one of x_1 and x_2 is on the cycle, then the claim is proved. Otherwise, the line segment $x_1 x_2$ crosses some edge $u_i u_{i+1}$ of the cycle. By the triangle inequality,

$$d_E(x_1, u_i) + d_E(x_2, u_{i+1}) \leq d_E(x_1, x_2) + d_E(u_i, u_{i+1}) \leq r_{2t} + r_t/2 \leq 2r_{2t},$$

which implies that at least one of $x_1 u_i$ and $x_2 u_{i+1}$ is an edge in R_{2t} . Therefore, there exists a path in R_{2t} from v to at least one of u_i and u_{i+1} , proving the claim. \square

The following theorem about the size of the giant component follows directly from Theorem 10.9 and Proposition 9.21 in [26].

Theorem 2.7 ([26]). *There exists a constant a_{big} so that a.s. a random geometric graph $RGG(N, r)$ with $\pi N r^2 > a_{\text{big}}$ has a connected component containing at least $0.99N$ of its vertices.*

Theorem 1.1 now follows directly from the following theorem.

Theorem 2.8. *Let $G = G_n$ be a graph produced by the SPA model with parameters $A_1 \in [0, 1)$, $A_2 > \max\{a_{\text{big}}, a_{\text{large}}\}$, $m = 2$, $p = 1$. Then a.s. R_n has a giant component C_n which contains at least $\sqrt{0.9n}$ vertices, and the diameter of C_n is $O(\log^2 n)$. Therefore, the effective diameter of G_n is $O(\log^2 n)$.*

Proof. Let k be the smallest integer so that $n2^{-k} \leq \log n$. On the one hand, by Lemma 2.5, a.s. we have that all random geometric graphs $R_n, R_{n/2}, \dots, R_{n2^{-k}}$ are nice. Let C_n denote the spanning component of R_n . Let y_0 be a vertex in C_n . By repeated application of Lemma 2.6, a.s. there exists a sequence of vertices y_1, y_2, \dots, y_k , with the following properties: for each i , $0 < i \leq k$, y_i is in the spanning component of $R_{n2^{-i}}$, and there exists a path of length $O(\log n)$ from y_i to y_{i-1} . Now $R_{n2^{-k}}$ has size at most $\log n$. Therefore, there exists a path of length $O(\log^2 n)$ between any two vertices in C_n .

On the other hand, by Theorem 2.7 and the fact that R_n is distributed as $RGG(n, r_n)$ with $r_n = \sqrt{\frac{A_2}{\pi n}}$, we have that a.s. R_n contains a giant component which contains at least $0.99n > \sqrt{0.9n}$ vertices. It is easy to see that this giant component must be the same as the spanning component C_n , and this completes the proof. \square

The methods used in this section do not suffice to show that the diameter of the giant component of G_n is also logarithmic. In principle, it could be that there exist vertices in G_n that are not contained in the giant component of R_n , but that are connected to this component by a long path that uses edges from inside the minor components of the graphs $R_n, R_{n/2}$, etc. Nevertheless, we believe that the SPA model graphs have logarithmic diameter inside their giant components a.a.s., and we leave this as an open problem.

3. LOWER BOUNDS FOR RUMOUR SPREADING

In this section we prove Theorem 1.4. Recall that m denotes the dimension and n denotes the number of vertices. We will first establish some structural properties of the graph generated by the SPA model, and then use these to prove results about the rumour spreading protocols. Let c_m denote the volume of the m -dimensional ball of unit radius. The proof is based on a classification of edges according to their length, and vertices and edges according to their birth time.

Definition 3.1. Let $\tau = \tau(n) = n^\beta$ and $L = L(n) = n^{-\eta}$, where $\eta, \beta \in (0, 1)$ and

$$\eta m < \beta(1 - pA_1). \quad (2)$$

Say an edge is *long* if the distance between its endpoints is larger than L , and is *short* otherwise. A vertex/edge is *old* if it was born during one of the rounds $1, 2, \dots, \tau$, and is *new* otherwise.

The following lemma establishes properties of old and new vertices and long and short edges.

Lemma 3.2 (Structural properties of the SPA model). *Let G be a graph generated by the SPA model with $pA_1 < 1$. Let $\varepsilon \in (0, 1)$ be a constant independent of n . Let $\tau = n^\beta$ and $L = n^{-\eta}$ be as in Definition 3.1, and let $y = y(n)$ be a function satisfying*

$$y = n^\gamma, \text{ for some } \gamma \in (\beta, 1 - \varepsilon/pA_1) \quad (3)$$

A.a.s. we have the following properties.

- (a) *All new edges are short.*
- (b) *If β and γ are such that*

$$\beta < (\gamma - \beta)(pA_1)/2, \quad (4)$$

then all old vertices v satisfy

$$\frac{\deg(v, \tau)}{\deg(v, n)} < n^\varepsilon \left(\frac{y \log n}{n} \right)^{pA_1}. \quad (5)$$

By part (a) of the lemma, all edges created after round τ are short, and thus the left hand side of (5) gives an upper bound on the proportion of edges incident with vertex v that are long. By (3), the right hand side of (5) is $o(1)$. Therefore, assertion (5) quantifies the informal statement “most edges incident to an old vertex are short.” The proof of the above lemma is somewhat technical and can be found at the end of this section.

Theorem 3.3 (Main Theorem for rumour spreading). *Let $\tau = n^\beta$ and $L = n^{-\eta}$ be as in Definition 3.1. Let $\varepsilon \in (0, 1)$, and let $y = n^\gamma$ be a function satisfying (3) and (4). Let $T = T(n) = n^\alpha$, where α is such that*

$$\alpha + \beta + (\varepsilon - pA_1) + \gamma pA_1 < 0. \quad (6)$$

Then, if the rumour starts from a uniformly random vertex, a.a.s. after T rounds of the push&pull protocol, all informed vertices lie within distance TL of the initial vertex.

Proof. Let G be a graph generated by the SPA model, and let ζ be a uniformly random vertex of it. We may assume that G satisfies the properties (a) and (b) given in Lemma 3.2. Note first that, if the rumour passes only through short edges, then in each time step the rumour can only spread to vertices that are within distance L of any vertex with the rumour. Thus, in T time steps the rumour can only reach vertices within distance TL of the initial vertex.

Let B denote the bad event “the rumour passes through a long edge during the first T rounds.” We need only show that a.a.s. B does not happen. Note that new vertices are not incident to long edges by Lemma 3.2(a). Moreover, by Lemma 3.2(b) every old vertex v satisfies (5), which guarantees that most edges incident to v are short. Condition (5) implies that the probability that an old vertex contacts a neighbour along some long edge in a given round is smaller than $n^\varepsilon \left(\frac{y \log n}{n}\right)^{pA_1}$. There are exactly τ old vertices. By the union bound over all old vertices and over the rounds 1 to T , we find that

$$\mathbb{P}[B] \leq \tau T n^\varepsilon \left(\frac{y \log n}{n}\right)^{pA_1} = n^{\beta + \alpha + \varepsilon + (\gamma - 1)pA_1} \log^{pA_1} n = o(1) \quad \text{by (6).} \quad \square$$

Proof of Theorem 1.4. Set

$$\begin{aligned} \delta &:= a(1 - a)/K - \alpha > 0, \\ \tau &:= n^{ma/K}, \\ y &:= n^{m(2+a)/K + \delta}, \\ L &:= n^{-a(1-a)/K + \delta/2}, \text{ and} \\ T &:= n^\alpha = n^{a(1-a)/K - \delta}, \\ \varepsilon &:= \min\{am/K, \delta(1 + a)\}/2, \end{aligned}$$

and observe that (2), (3), (4) and (6) are satisfied, and that $TL = o(1)$. By Theorem 3.3, a.a.s. after T rounds of the push&pull protocol, all informed vertices lie in a ball of volume $O((TL)^m) = o(1)$. By a standard Chernoff bound, a.a.s. the number of vertices in any such ball is $O(n(TL)^m) = o(n)$. \square

In the rest of this section we prove Lemma 3.2. We will use the following two concentration bounds, sometimes called multiplicative Chernoff bounds (see, e.g., [24, Theorem 2.3(b,c)]). Let X be a sum of independent indicator random variables and let $\delta \geq 0$. We have

$$\mathbb{P}[X \leq (1 - \delta)\mathbb{E}[X]] \leq \exp(-\delta^2 \mathbb{E}[X]/2) \quad (7)$$

and

$$\mathbb{P}[X \geq (1 + \delta)\mathbb{E}[X]] \leq \exp\left(-\frac{\delta^2 \mathbb{E}[X]}{2 + 2\delta/3}\right). \quad (8)$$

We will use the following theorem from [19].

Theorem 3.4 (Theorem 5.2 in [19]). *Let $f(n)$ be any function tending to infinity with n . Let v be a vertex with $\deg^-(v, R) \geq f(n) \log n$. Then, with probability at least $1 - O(n^{-4/3})$, for all $r \in \{R, R+1, \dots, 2R\}$ we have*

$$|\deg^-(v, r) - \deg^-(v, R)(r/R)^{pA_1}| \leq \frac{2r\sqrt{\deg^-(v, R) \log n}}{pA_1 R}.$$

In particular, setting $r = 2R$, the above theorem implies that if $\deg^-(v, R) \geq f(n) \log n$ for some $f(n) = \omega(1)$, then

$$\mathbb{P}[\deg^-(v, 2R) \geq (2^{pA_1} - o(1)) \deg^-(v, R)] \geq 1 - O(n^{-4/3}). \quad (9)$$

Lemma 3.5. *Let $\delta, \varepsilon \in (0, 1)$ be arbitrary constants, and let $\tau = n^\beta$ and $y = n^\gamma$ be functions satisfying (3). For any old vertex v we have*

$$\mathbb{P}\left[\frac{\deg^-(v, \tau)}{\deg^-(v, n)} \geq n^\varepsilon \left(\frac{y \log n}{n}\right)^{pA_1}\right] = O\left((\tau/y)^{(1+o(1))\delta^2 pA_1/2} + \frac{\log n}{n^{4/3}}\right).$$

Proof. If $d := \deg^-(v, \tau) = 0$, the conclusion is obvious, so assume that $d \geq 1$. Define the following events:

$$\begin{aligned} E_1 &:= \{\deg^-(v, y) \geq (1 - \delta)pA_1 d \log(y/\tau)\}, \\ E_2 &:= \{\deg^-(v, y \log n) \geq (1 - \delta)^2(pA_1)^2 d \log(y/\tau)(\log \log n)\}, \text{ and} \\ E_3 &:= \{\deg^-(v, n) \geq (1 - \delta)^2(pA_1)^2 d \log(y/\tau)(\log \log n)(n/(y \log n))^{pA_1} n^{-\varepsilon}\}. \end{aligned}$$

We remark that $\log(y/\tau) = (\gamma - \beta) \log n$. Note that if E_3 happens then

$$\frac{d}{\deg^-(v, n)} < \left(\frac{y \log n}{n}\right)^{pA_1} n^\varepsilon.$$

Hence to prove the lemma we want to bound the probability of E_3^c . We will prove that the probabilities $\mathbb{P}[E_1^c]$ and $\mathbb{P}[E_2^c|E_1]$ are at most $O((\tau/y)^{(1+o(1))\delta^2 pA_1/2})$, and that $\mathbb{P}[E_3^c|E_1, E_2]$ is at most $O(n^{-4/3} \log n)$. This would prove the lemma, since

$$\mathbb{P}[E_3^c] \leq \mathbb{P}[E_1^c] + \mathbb{P}[E_3^c|E_1] \leq \mathbb{P}[E_1^c] + \mathbb{P}[E_2^c|E_1] + \mathbb{P}[E_3^c|E_1, E_2].$$

First, we bound $\mathbb{P}[E_1^c]$. Note that for each $i \in \{\tau+1, \dots, y\}$, the probability that v_i creates an edge to v is at least $pA_1 d/i$. In fact, $\deg^-(v, y) - \deg^-(v, \tau)$ is stochastically larger than the sum of $y - \tau$ independent indicator variables $X_{\tau+1}, \dots, X_y$ with $\mathbb{E}[X_i] = pA_1 d/i$, as in this formula we have ignored the neighbours accumulated in rounds $\tau+1, \dots, y$. Then

$$\mathbb{E}\left[\sum X_i\right] = (1 + o(1))pA_1 d(\log(y/\tau)),$$

hence by the multiplicative Chernoff bound (7),

$$\mathbb{P}[E_1^c] \leq \exp(-\delta^2(1 + o(1))pA_1 d \log(y/\tau)/2) = (\tau/y)^{(1+o(1))pA_1 d \delta^2/2}.$$

Second, we bound $\mathbb{P}[E_2^c|E_1]$. Conditional on E_1 , by a similar argument, the difference in in-degrees $\deg^-(v, y \log n) - \deg^-(v, y)$ is stochastically larger than the sum of $y \log n - y$

independent indicator variables $Y_{y+1}, \dots, Y_{y \log n}$ with $\mathbb{E}[Y_i] = (1 - \delta)(pA_1)^2 d \log(y/\tau)/i$. Since

$$\mathbb{E} \left[\sum Y_i \right] = (1 + o(1))(1 - \delta)(pA_1)^2 d \log(y/\tau)(\log \log n),$$

by the multiplicative Chernoff bound (7),

$$\mathbb{P}[E_2^c | E_1] \leq \exp(-(1 + o(1))\delta^2(1 - \delta)(pA_1)^2 d \log(y/\tau)(\log \log n)/2) = (\tau/y)^{-\Omega(d \log \log n)}.$$

Finally, conditional on E_1 and E_2 we may use (9) repeatedly for $R = y \log n, 2y \log n, \dots$, all the way up to $R = n/2$ to obtain that with probability at least $1 - O((\log n)n^{-4/3})$ we have

$$\deg^-(v, n) \geq (2^{pA_1} - o(1))^{\log_2(n/(y \log n))} \deg^-(v, y \log n) \quad (10)$$

$$> (n/(y \log n))^{pA_1} n^{-\varepsilon} \deg^-(v, y \log n), \quad (11)$$

completing the proof. Note that condition (3) together with E_2 ensure that $\deg^-(v, y \log n) = \omega(\log n)$ hence (9) can indeed be applied. \square

The following result follows from the proof of [1, Theorem 1.5].

Theorem 3.6. *A.a.s. all vertices have outdegree $O(\log^2 n)$.*

Since ε can be chosen arbitrarily in Lemma 3.5, and all outdegrees are polylogarithmic a.a.s., we may replace indegrees with total degrees in Lemma 3.5 and conclude the following.

Corollary 3.7. *Let $\tau = \tau(n) = n^\beta$ and $y = y(n) = n^\gamma$ be functions satisfying (3), and let $\delta, \varepsilon \in (0, 1)$ be arbitrary constants. For any old vertex v we have*

$$\mathbb{P} \left[\frac{\deg(v, \tau)}{\deg(v, n)} \geq n^\varepsilon \left(\frac{y \log n}{n} \right)^{pA_1} \right] = O \left((\tau/y)^{(1+o(1))\delta^2 pA_1/2} + \frac{\log n}{n^{4/3}} \right).$$

We will use the following theorem from [8].

Theorem 3.8 (Theorem 2.3 in [8]). *Let $f(n)$ be any function that goes to infinity with n . A.a.s. for all $i \in \{1, 2, \dots, n\}$ and all $t \in \{i, \dots, n\}$ we have*

$$\deg^-(v_i, t) = O(f(n)(\log n)(t/i)^{pA_1}).$$

We now have all the ingredients to prove Lemma 3.2.

Proof of Lemma 3.2. (a) By Theorem 3.8, a.a.s. for all $i \in [n]$ and all $t \in \{i, \dots, n\}$ we have

$$\deg^-(v_i, t) \leq (\log n)^{3/2} (t/i)^{pA_1}.$$

Suppose this is the case and let $i \in [n]$. Then at any time $t \in \{\max\{i, \tau + 1\}, \dots, n\}$, the sphere of influence of v_i at time t has volume at most

$$\begin{aligned} \frac{A_1(\log n)^{3/2} (t/i)^{pA_1} + A_2}{t} &\leq A_1(\log n)^{3/2} t^{pA_1-1} + \frac{A_2}{t} \\ &< \tau^{pA_1-1} \log^2 n && \text{since } pA_1 < 1 \\ &< c_m L^m && \text{by (2)} \end{aligned}$$

Thus any incoming edge to v_i that is created after round w is short.

(b) The number of old vertices is $\tau = n^\beta$, and the probability that an old vertex fails to satisfy (5) is $O((\tau/y)^{(1+o(1))\delta^2 p A_1/2}) + O(n^{-4/3} \log n)$ for any constant $\delta \in (0, 1)$ by Corollary 3.7. Recall that $(\tau/y) = n^{-(\gamma-\beta)}$, where $\beta < \gamma$. By (4) we can choose δ close enough to 1 so that $(\gamma - \beta)(1 + o(1))\delta^2 p A_1/2 > \beta$. The union bound completes the proof. \square

REFERENCES

- [1] W. Aiello, A. Bonato, C. Cooper, J. Janssen, and P. Prałat. A spatial web graph model with local influence regions. *Internet Math.*, 5(1-2):175–196, 2008.
- [2] A. Bonato, D. Gleich, D. Mitsche, P. Prałat, Y. Tian, and D. Young. Dimensionality of social networks using motifs and eigenvalues. *PLoS ONE*, 9(9): e106052, 2014.
- [3] A. Bonato, J. Janssen, and P. Prałat. Geometric protean graphs. *Internet Mathematics*, 8(1-2):2–28, 2012.
- [4] A. Bonato, J. Janssen, and E. Roshanbin. Burning a graph as a model of social contagion. In Anthony Bonato, Fan Chung Graham, and Paweł Prałat, editors, *Algorithms and Models for the Web Graph*, volume 8882 of *Lecture Notes in Computer Science*, pages 13–22. Springer International Publishing, 2014.
- [5] M. Bradonjić, A. Hagberg, and A. Percus. The structure of geographical threshold graphs. *Internet Math.*, 5:113–140, 2008.
- [6] K. Bringmann, R. Keusch, and J. Lengler. Average distance in a general class of scale-free networks with underlying geometry. 2016. Preprint, available at <http://arxiv.org/abs/1602.05712>.
- [7] C. Cooper and A. Frieze. The cover time of the preferential attachment graph. *J. Combin. Th. B*, 97:269–290, 2007.
- [8] C. Cooper, A. Frieze, and P. Prałat. Some typical properties of the spatial preferred attachment model. *Internet Mathematics*, 10(1-2):116–136, 2014.
- [9] A. Demers, D. Greene, C. Hauser, W. Irish, J. Larson, S. Shenker, H. Sturgis, D. Swinehart, and D. Terry. Epidemic algorithms for replicated database maintenance. In *Proceedings of PODC '87*, pages 1–12, New York, NY, USA, 1987. ACM.
- [10] B. Doerr, M. Fouz, and T. Friedrich. Social networks spread rumors in sublogarithmic time. In *STOC'11—Proceedings of the 43rd ACM Symposium on Theory of Computing*, pages 21–30. ACM, New York, 2011.
- [11] U. Feige, D. Peleg, P. Raghavan, and E. Upfal. Randomized broadcast in networks. *Random Struct. Algorithms*, 1(4):447–460, 1990.
- [12] A. Flaxman, A. Frieze, and J. Vera. A geometric preferential attachment model of networks. II. *Internet Math.*, 4(1):87–111, 2007.
- [13] N. Fountoulakis, K. Panagiotou, and T. Sauerwald. Ultra-fast rumor spreading in social networks. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1642–1660. ACM, New York, 2012.
- [14] T. Friedrich, T. Sauerwald, and A. Stauffer. Diameter and broadcast time of random geometric graphs in arbitrary dimensions. *Algorithmica*, 67(1):65–88, 2013.
- [15] A. M. Frieze and G. R. Grimmett. The shortest-path problem for graphs with random arc-lengths. *Discrete Appl. Math.*, 10(1):57–77, 1985.
- [16] G. Ganesan. Size of the giant component in a random geometric graph. *Annales de l'Institut Henri Poincaré*, 49:1130–1140, 2013.
- [17] E. Jacob and P. Mörters. Robustness of scale-free spatial networks. 2015. Preprint, available at <http://arxiv.org/abs/1504.00618>.
- [18] E. Jacob and P. Mörters. Spatial preferential attachment networks: Power laws and clustering coefficients. *Annals of Applied Probability*, 25(2):632–662, 2015.
- [19] J. Janssen, P. Prałat, and R. Wilson. Geometric graph properties of the spatial preferred attachment model. *Advances in Applied Mathematics*, 50(2):243–267, 2013.
- [20] J. Janssen, P. Prałat, and R. Wilson. Nonuniform distribution of nodes in the spatial preferential attachment model. *Internet Mathematics*, 12(1-2):121–144, 2016. Available at [arXiv:1506.06053](https://arxiv.org/abs/1506.06053).

- [21] R. Karp, C. Schindelhauer, S. Shenker, and B. Vöcking. Randomized rumor spreading. In *41st Annual Symposium on Foundations of Computer Science (FOCS'00)*, pages 565–574. IEEE Comput. Soc. Press, Los Alamitos, CA, 2000.
- [22] D. Kempe, J. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '03, pages 137–146, 2003.
- [23] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graph evolution: Densification and shrinking diameters. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(1), March 2007. Article 2.
- [24] C. McDiarmid. Concentration. In *Probabilistic methods for algorithmic discrete mathematics*, volume 16 of *Algorithms Combin.*, pages 195–248. Springer, Berlin, 1998.
- [25] R. Pastor-Satorras and A. Vespignani. Epidemic dynamics and endemic states in complex networks. *Phys. Rev. E*, 63:066117, May 2001.
- [26] M. Penrose. *Random geometric graphs*. Oxford Studies in Probability. Oxford Un. Press, 2003.
- [27] K. Zuev, M. Boguñá, G. Bianconi, and D. Krioukov. Emergence of soft communities from geometric preferential attachment. *Nature Scientific Reports*, 5:9421, 2015.

JJ: DEPARTMENT OF MATHEMATICS AND STATISTICS, DALHOUSIE UNIVERSITY, 6316 COBURG ROAD,
 PO BOX 15000, HALIFAX, NOVA SCOTIA, CANADA B3H 4R2
E-mail address: jeannette.janssen@dal.ca

AM: UNIVERSITY OF WATERLOO, WATERLOO, ON, CANADA
Current address: Department of Computer Science, University of British Columbia, 2366 Main Mall,
 Vancouver, B.C., V6T 1Z4 Canada
E-mail address: AbbasMehrabi@gmail.com
URL: <http://www.cs.ubc.ca/~amehrabi/>